The Dispersion of Neuronal Clones Across the Cerebral Cortex

C. Walsh and C. L. Cepko (*1*) recently reported that clonally related neurons are found in widely dispersed areas of rat cerebral cortex. They used a "cocktail" of 100 identifiable retroviruses as genetic tags to label cells in brains that were infected during neurogenesis in the embryo. Brains were removed three or more days after birth, and clonal analysis was performed by amplifying the genetic tags with polymerase chain reaction. Several widely dispersed repetitions of individual tags were found, and on the basis of low estimates for the probabilities of multiple infections with the same tag, Walsh and Cepko concluded that these were most likely to be clones of cells that had become widely separated during neurogenesis. The statistical method used by Walsh and Cepko to interpret their results is based on the assumption that all 100 tags were equally represented in the material used for the injections. This assumption appears to be unjustified, and the probability calculations are therefore questionable.

Equal titer of all 100 tags would be difficult to achieve for several reasons. Titration experiments produce variable results, and the titers of many different viruses must be compared. Furthermore, 100 viruses (*1*) were raised in pairs from 50 producer lines, so independent titration of viruses within pairs was not possible. As evidence for the titers being equal, Walsh and Cepko present figure 5, which shows the number of times each of the 100 tags was recovered in a total of 235 clones and which comprises pooled data from 16 experiments. Twenty-one tags showed no recoveries.

We simulated the drawing of 235 clones from an equal mixture of 100 tags. Only three out of 10^5 draws resulted in 21 or more tags showing no recovery. Walsh and Cepko's data are therefore inconsistent with equal representation of all tags in the mixture, so the original mixture was apparently nonuniform.

We then estimated how much variability in the original cocktail was consistent with the data presented in (*1*). We assumed (arbitrarily) a log normal distribution for the concentration of each tag with a mean of 1%. We then varied the standard deviation (SD) until the expected number of tags showing no recovery was about 21, as Walsh and Cepko observed. A value for SD (in natural logarithm units) of 0.7 resulted in an average of 20.9 tags showing no recovery in 10^6 random draws of 235 clones. With this distribution, there was extensive variation in the concentrations of individual tags, such that a third of the tags had either less than half or more than twice the nominal 1% distribution assumed by Walsh and Cepko.

Introducing this distribution in the Monte Carlo approach used by Walsh and Cepko, we calculated the probabilities of independently drawing the same tag twice or more within a brain. Ten thousand simulations were run for each of three values of k, the number of clones per brain (Table 1). The probabilities of double and triple occurrences of the same tag were higher if the nonuniform distribution of tags was taken into account. Similar results were obtained when we assumed other forms of distribution, such as a normal distribution, for the tag concentrations in the original mixture.

We also calculated the probabilities of drawing two different tags three times in the same brain in order to assess the result described in figure 2 of (*1*). With the use of Walsh and Cepko's assumption of a uniform distribution, we confirmed their probability of 0.0006 (our simulations produced a figure of 0.0007). With the use of our method, we obtained a probability of 0.005. We therefore found that the probability in a single experiment of obtaining by chance the result shown in figure 2 of (*1*) is 0.5% (1 in 200), nearly tenfold more than the 0.06% (1 in 1670) stated in (*1*). As the number of experiments in the series increases so would the probability of finding such a striking result.

There is another potential source of error not considered by Walsh and Cepko or examined by us in any detail because of lack of data. The data in figure 5 of (*1*) are an aggregate of results from several experiments. Any inequality between experiments that results from the use of different stocks or serial dilutions would have been averaged out by such aggregation. Therefore, the probabilities we have calculated may have been underestimated.

In summary, we calculate that the incidence of multiple infection will be greater than estimated in (*1*); most brains will have double hits, and even more extreme results will occur regularly by chance. Unless the extent of the inequality of viral titers is known, however, the true incidence of multiple infections cannot be calculated. Independent evidence is required in a study such as (*1*) to establish precisely the viral titers of the injected material or to show that multiple hits do not occur.

Thomas B. L. Kirkwood
Jack Price
Elizabeth A. Grove
National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, United Kingdom

Table 1. Probabilities (frequencies) of observing double and triple occurrences of the same tag in a brain on the basis of 10000 Monte Carlo trials. k indicates the number of clones found per brain. ER indicates results using the method and assumption of equal representation of tags in (*1*). NU indicates results using a nonuniform distribution of tag concentrations.

Occurrences of drawing the same tag	Total clones k	Probability	
		ER	NU
Double	10	0.37	0.52
Double	15	0.67	0.82
Double	20	0.86	0.96
Triple	10	0.01	0.04
Triple	15	0.04	0.13
Triple	20	0.10	0.27

REFERENCES

1. C. Walsh and C. L. Cepko, *Science* **255**, 434 (1992).

20 April 1992; accepted 31 August 1992

Response: We welcome the opportunity to discuss statistical considerations at more length than could be accommodated in our original study of cell lineage in the cerebral cortex (*1*).

In our study, cell lineage in the cortex was determined by marking clones of cortical cells with a retroviral library. The library was constructed to contain approximately 100 members that each carried a DNA insert as a genetic "tag." Cells belonging to a given clone will contain the same tag regardless of where the cells migrate. However, two or more different clones that coincidentally carry the same tag will produce a false appearance of one widespread clone. For each injected cortical hemisphere ("experiment"), the probabilities of coincidental double, triple, quadruple, and so forth, infections were calculated, and compared to the frequency of the widespread appearances of a single tag. A computer simulation calculated the probability of coincidental infections by picking t elements from a mixture of 100 equally represented tags (where t equals the numbers of distinct tags seen in the experiment plus the number of widespread subunits of the clone under consideration minus one). Thus, in an experiment with five tags, if one tag were in three widespread cells, the simulation calculated the probability that one tag would be chosen three times when seven (five plus three minus one) elements are picked, with replacement, from 100

tags. This simulation was repeated 10,000 times, and the frequency of coincidental infections by the same tag was determined and used as a probability. In the most heavily infected brains, widespread appearances of the same viral tag in two cells had a high probability of being the result of coincidental double infections. However, widespread patterns were frequent even in lightly infected brains, and some widespread patterns included three, four, five, or even seven widely dispersed cells. Statistical analysis of widespread patterns containing more than two cells suggested that at least some represented widespread dispersion of cells derived from a common progenitor.

Kirkwood *et al.* analyzed our empirical data on the recovery of the tags in the library, shown in our figure 5 (*1*), and recognized that the number of different tags recovered in our experiments provides a measure of the complexity of the library sampled. As only 79 of 100 tags were recovered in 235 clones, their analysis demonstrated that it was unlikely that the 100 tags were equally represented in the viral mixture. Our own subsequent calculations verified this conclusion. We disagree, however, with their assertion that the content of the library implied by this data has a significant impact on the probability that widespread clones occur. The frequency of widespread clones is high enough that the phenomenon could not be due to chance occurrence, regardless of whether the library contained 79 or 100 members and whether tags were distributed equally or nonuniformly.

In trying to arrive at a better mathematical model to describe the empirically determined ratio of tags, Kirkwood *et al.* assumed that concentrations of the 100 tags were represented by a log normal distribution. They varied the standard deviation of this distribution such that 79 of 100 tags would be recovered in 235 clones. We compared the expected patterns of tag recovery using the model of Kirkwood *et al.* to the expected pattern of recovery of 100 tags present in equal ratios (Table 1). The bottom row of Table 1 indicates how closely the predicted values match the number of times each tag was recovered (rather than only analyzing tags that were not recovered, as Kirkwood *et al.* did), for each model. The model of Kirkwood *et al.* gives a better fit than our original assumption of 100 equally represented tags. Other models, however, can be derived that are still more accurate.

A likely possibility is that many of the 100 tags are represented fairly equally, while others of the 100 tags were not recovered for a variety of technical reasons. For example, the DNA used to prepare the retroviral supernatants may have been of poor quality or insufficient quantity, so that little or no retrovirus carrying that tag was produced (*2*). Alternatively, some DNA sequences might have been lost when they were put into retroviral vectors. We therefore deliberately designed the library to greatly exceed the complexity needed for statistical significance to ensure that, if some tags were lost, the data would still be easily interpretable. The expected patterns of tag recovery for mixtures of 81 or 85 equally represented tags (Table 1) match the data of our original figure 5 better than the distribution used in Kirkwood *et al.*'s model.

Table 1. Comparison of statistical models of tag distribution and empirically determined tag distributions. Five models of tag distributions are analyzed and compared with the observed appearances of viral tags illustrated in figure 5 of (*1*). For each model, the total error [sum (| observed-calculated |) and the sum of the squared errors (sum (| O-C |^2) are shown in the bottom rows. Smaller numbers indicate closer fits to the observed data. The column designated Kirkwood *et al.* assumes 100 tags distributed in log normal fashion with an SD of 0.7 natural logarithm units. The next three columns show expected patterns assuming 100, 85, or 81 tags distributed uniformly (SD = 0). The last column shows expected patterns of a combined model, derived to minimize the total error. This distribution was arrived at as described (*3*) and consists of 86 tags, with log normal distributions, SD = 0.36 natural logarithm units.

Clones (No.)	Number of tags					
	Observed	Kirkwood *et al.*	100 tags	85 tags	81 tags	Combined
0	21	20	9	20	24	22
1	18	25	23	15	13	16
2	17	19	27	20	19	19
3	15	13	21	19	18	16
4	14	8	12	13	13	11
5	8	5	5	7	8	7
6	7	3	2	3	4	4
>6	0	7	1	2	2	4
Sum (| observed-calculated |)		29	42	19	19	17
Sum (|O-C |^2)		168	344	57	61	45

Finally, we used a model combining the log normal distribution of Kirkwood *et al.* with the idea that some tags were not represented at all in order to achieve the best possible fit to the empirical data with the use of log normal tag distributions. A computer simulation (*3*) varied the number of tags and the standard deviation of their log normal distribution so as to minimize the total error between the expected tag distribution and the empirically observed tag distribution. Results with this "combined" model, shown in the last column of Table 1, provide a slightly closer match to the empirical data.

All the different models give largely similar results when used for statistical analysis. In order to illustrate the differences in models, one of us (G.M.C.) wrote computer programs that generate analyses using each of the models described above. We then re-analyzed data from ten experiments reported in our article (*1*), using four models: (i) Kirkwood *et al.*'s simulation, (ii) assuming that 80 tags are equally represented, (iii) the original assumption that 100 tags are equally represented, and (iv) the combined model described above. When Kirkwood *et al.*'s model was used, seven of these ten experiments showed widespread patterns for which the probability (*P*) is less than or equal to 0.04 that the patterns are coincidences. The likelihood that an event with such a probability will occur by chance in seven or more of ten trials ($P^{7/10}$) can be calculated using a binomial equation with $P = 0.04$

$$P^{7/10} = \frac{10!}{7!3!}(0.04)^7(0.96)^3 + \frac{10!}{8!2!}(0.04)^8(0.96)^2 + \ldots + \frac{10!}{10!}(0.04)^{10}$$

$$P^{7/10} < 0.00000002$$

Therefore, the results of Kirkwood *et al.*'s statistical analysis strongly confirm our own analysis, rather than calling it into question. Our original assumption of 100 equally represented tags, or any of the simulations except Kirkwood *et al.*'s, confirms that eight of ten experiments showed events with a probability less than or equal to 0.04. The probability of this occurring by chance is less than 0.0000000003. We have tried other statistical models as well (*4*), and these also produced statistically significant results. As can be seen from Table 2,

Table 2. Statistical analysis of widespread clones with the use of four statistical models. All widespread patterns were analyzed for ten experiments reported previously (*1*). Each row in the table shows one clone, defined by tags amplified by polymerase chain reaction; brains with more than one nonclustered clone have each clone on a separate row and are indicated as A, B, C, D, E, and F. For each widespread pattern, the probability that it is due to coincidental infections by one tag is listed. Each widespread pattern has subunits, so the probability of multiple infections is calculated by one tag in "*t*" hypothetical clones (defined in the text) (*11*); ns, not significant ($P > 0.05$).

Experiment (No. of tags)	Clone	Clusters (No.)	*t**	*P*			
				Kirkwood *et al.*	80	100	Combined
E15-12, left (4 tags)	A	2	5	ns	ns	ns	ns
	B	4	7	0.001	0.0002	<.0001	<.0001
	C	7	10	<.0001	<.00001	<<.00001	<.0001
E15-2, right (6 tags)	A	2	7	ns	ns	ns	ns
	B	3	8	0.019	<.01	<.01	0.012
E15-4, left (8 tags)	A	3	10	0.04	0.02	0.01	0.022
E15-13, right (9 tags)	A	2	10	ns	ns	ns	ns
	B	3	11	0.077	0.03	0.03	0.030
E15-4, right (11 tags)	A	3	13	0.09†	0.04	0.03	0.050
	B	3	13	0.09†	0.04	0.03	0.050
E15-9, left (15 tags)	A	2	16	ns	ns	ns	ns
	B	2	16	ns	ns	ns	ns
	C	4	18	0.02	0.005	0.003	0.007
E15-8, right (17 tags)	A	2	18	ns	ns	ns	ns
	B	2	18	ns	ns	ns	ns
	C	2	18	ns	ns	ns	ns
	D	2	18	ns	ns	ns	ns
	E	2	18	ns	ns	ns	ns
	F	5	21	0.009	0.0005	0.0004	0.001
E17-4, left (2 tags)	A	2	3	0.038	0.04	0.03	0.040
E17-5, left (3 tags)	A	2	4	ns	ns	ns	ns
E17-3, right (13 tags)	A	2	14	ns	ns	ns	ns

*Total hits in the brain assuming that the widespread clone results from coincidental hits by one tag, t = (number of tags) + (number of clusters − 1). †$P = 0.005$ that both events occur in the same experiment.

and as we emphasized in the text and in figure 6 of our article, we illustrated typical data, not the most extreme patterns. Several other patterns, some illustrated in figure 6 of our article, were statistically far less likely to have represented coincidences than the patterns illustrated in our figure 2.

Thus, with any statistical model we have tried, all of the conclusions stand: most important, (i) some cortical clones disperse very widely, crossing functional boundaries and cytoarchitectonic areas; and (ii) clonal analysis with a single retroviral marker causes unpredictable errors because of widespread clonal dispersion (*5*). The precise number and types of errors caused by using a single viral marker depend on the assumptions made and the statistics used. However, these errors are always a problem with single marker studies. Inattention to these errors obscures the biology of the system.

Kirkwood *et al.* suggest that another potential source of sampling error could come from the dilution of viral stocks. As each stock was diluted at most a single time, and each aliquot contained a high number of viral particles, stock preparation did not add significant nonuniformities that would affect the probability calculations (*6*).

There are additional reasons that make it unlikely that widespread clones are spurious. Coincidental double infections should be more common in experiments in which larger numbers of clones were labeled, but this was not the case. Widespread clones also occurred in some reproducible patterns, suggesting that they reflect biological phenomena and not random coincidences.

We have subsequently done more experiments, using the same viral library, with less than five clones per brain, where the probability of double infections by the same tag is low (Table 2), and confirmed our previous finding (*7*).

Widespread dispersion of fluorescently labeled cells in the ventricular zone of the developing cortex has been recently observed (*8*), as has migration of some cells beneath the cortex in the transverse plane (*9*), which would also be expected to produce widespread clonal dispersion. Results from these nonretroviral studies provide independent confirmation of retrovirally derived results and suggest mechanisms to explain them.

With several converging lines of evidence supporting the same conclusion, and with the statistical analysis of Kirkwood *et al.* providing firm support, we hope that this discussion lays to rest any doubts about our study. We encourage anyone who is interested in using retroviral libraries to contact us for samples, technical advice, or both (*10*).

Christopher Walsh
Department of Genetics, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115, and Department of Neurology, Massachusetts General Hospital, Fruit Street, Boston, MA 02114
Constance L. Cepko
Elizabeth F. Ryder
Department of Genetics, Harvard Medical School
George M. Church
Department of Genetics, Harvard Medical School, and Howard Hughes Medical Institute
Cliff Tabin
Department of Genetics, Harvard Medical School

REFERENCES AND NOTES

1. C. Walsh and C. L. Cepko, *Science* **255**, 434 (1992).
2. We have found one example of an unrecovered tag that was apparently lost because the DNA used to prepare the retroviral supernatant was of insufficient quantity.
3. The program calculated expected patterns of tag recovery and calculated the sum of errors between the calculated and observed tag ratios. The program first tested equal representations of 79 to 87 tags. Then it tested expected patterns of recovery for 81, 83, 85, and 100 tags, respectively, with the assumption that these tags were

distributed in lognormal patterns with an SD that was varied systematically in intervals of 0.05 from 0 to 0.9. Then it tested recovery patterns for 83, 84, 85, 86, 87, and 88 tags, respectively, testing SDs every 0.01 from 0.31 to 0.43. The smallest total error was obtained with 86 tags distributed with an SD equal to 0.36 natural logarithm units.

4. For example, we have also used the empirical concentration of each tag to calculate probabilities that the tag might be coincidentally present in multiple clones. This analysis shows, with similar statistical significance, that some clones are widely dispersed.
5. Clonal analysis with a single marker in some cases also falsely interprets cells that are part of two different clones as a single clone. These "lumping errors" are not affected by the choice of statistical model.
6. We used a series of viral stocks, each of which had been diluted once from a single concentrated stock. Each tube of virus that was used to infect one or more litters of animals was made by adding at least 3×10^5 colony-forming units (CFU) to varying volumes of diluent to give final dilutions of 1:5, 1:10, and 1:20. Three such tubes of diluted virus were used in all, and one animal was infected with undiluted virus stock directly pipetted from the concentrated stock. The titer of the concentrated virus stock was 3×10^7 CFU per milliliter, and thus the dilutions contained from 1.5×10^3 to 6×10^3 CFU per microliter. From each tube of diluted virus, a pipette was filled with 3 to 5 μl, and each injection was approximately 1 μl. Because of the large number of viral particles sampled in 1 μl of any dilution, nonuniformities introduced by sampling should have been minor. No new, additive error would have been introduced.
7. C. Walsh and C. L. Cepko, *Soc. Neurosci. Abstr.* **18**, 925 (1992). These experiments also provided an additional "control": when clones were labeled by infection at embryonic day 15 (E15) (as in our original study), but analyzed earlier, at E18, no widespread dispersion was seen in 12 clones. This would be expected if later widespread dispersion was caused by migration that had not yet had time to occur, but would not make sense if widespread dispersion resulted only from coincidental infections of different clones by the same tag. We have also made a new library with at least 150 tags (using alkaline phosphatase as a histochemical marker) and again observed many widely dispersed clones (C. Walsh and C. L. Cepko, unpublished results).
8. G. Fishell, C. A. Mason, M. E. Hatten, *Soc. Neurosci Abstr.* **18**, 926 (1992).
9. N. O'Rourke, M. E. Dailey, S. Smith, S. K. McConnell, *Science* **258**, 299 (1992).
10. The computer program for simulating statistical models (MONTAG), written by George Church (Department of Genetics, Harvard Medical School), is available through anonymous internet ftp from rascal.med.harvard.edu. It will run on most virtual memory operating system machines without recompiling. Type "run montag" and answer the queries. If there are problems, contact church@gnome.med.harvard.edu.
11. For example, in the experiment illustrated in figure 2A (*1*), three tags were present. One tag was present in a nonclustered clone with two subunits. Therefore, we calculated the probability that in three plus one equals four hypothetical clones, one tag is present coincidentally in two different clones. For the experiment shown in figure 2C (*1*), the two widespread clones (tags 47 and 52) were each simulated by calculating the probability of getting three hits by one tag out of 11 + 2 = 13 total hypothetical clones. These assumptions are conservative, because analyzing each clone separately understates how unlikely it is that multiple coincidences occur within one experiment.

12 June 1992; accepted 31 August 1992

"Several of the lab rats died last night. You haven't been feeding them your leftover lunch, have you?"

Response

Christopher Walsh, Constance L. Cepko, Elizabeth F. Ryder, George M. Church and Cliff Tabin

Science **258** (5080), 317-320.
DOI: 10.1126/science.258.5080.317-a